Guest Editors' Introduction

Design Challenges for High-Performance Network Interfaces

With the advent of distributed computing, developers are increasingly concerned with the efficient movement of data through the network and, in particular, the design of efficient network interfaces.

Andrew A. Chien
University of California, San Diego

Mark D. Hill
University of Wisconsin-Madison

Shubhendu S. Mukherjee
Compaq Computer Corporation

With the advent of distributed computing and the Internet, and with dramatically faster processor clock rates and complete systems on chips, computing is increasingly concerned with the efficient movement of data—across the interconnects within a machine, the system area network within a machine room, and the local area or wide area network. An increasingly critical issue in the design of computer systems is how to achieve efficient communication and, in particular, the design of network interfaces.

NETWORK INTERFACES

A network interface is a device that allows a computer to communicate with a network. Figure 1 shows a conventional network interface attached to an input/output (I/O) bus. Network interfaces can also be attached to memory controllers or even directly to the processor data path.

Network interface design has a crucial impact on communication efficiency—determining the cost of initiating and responding to communication actions, moving the data required for communication, and providing application isolation across communicating domains. The interaction overhead between a processor and a network interface is exacerbated by the increasing operation rates of microprocessors with gigahertz clocks and networks with gigabits-per-second bandwidth.

Two important components of communication are

- network software that manages the network and implements communication protocols, and
- network interface hardware that moves data, provides protection, and generates communication events.

However, these components manifest themselves in different forms, depending on the application, system context, and cost-performance requirements.

NETWORK CONNECTIVITY

Three major classes of network connectivity are

- workstations or PCs connected by LANs or WANs,
- workstations or PCs connected by a system-area network (SAN), and
- parallel processors connected by a custom network.

Traditionally, LANs and WANs have provided unreliable delivery, and thus computers connected this way use network software, such as TCP/IP protocol stacks, to ensure reliability. However, the cost of such protocol stacks is significant. Consequently, developing efficient and reliable network software has become a key research area. Performance optimizations include reducing the code path of protocol stacks and optimizing data movement between host memory and the network interface.

SANs—for example, Myricom's Myrinet and Compaq's Servernet—will soon deliver bandwidths of 10 Gbps and latencies of tens of nanoseconds to hosts in close proximity (say, 100 meters). This represents a two to four order-of-magnitude improvement over current LANs. These high performance levels and these networks' generally reliable delivery have inspired the use of lightweight protocols (such as Active Messages or Fast Messages) and innovative protection and notification structures. SAN-based computing provides a promising avenue for building large-scale systems (in computing, memory, and storage) using low-cost building blocks.

0018-9162/98/$10.00 © 1998 IEEE

Massively parallel processors (MPPs) are tightly integrated systems with the highest performance levels for communication and the deepest integration of such communication into the computing complex. This requirement is driven by fine-grain parallel applications that demand extremely low-latency and high-bandwidth communication. Primary research objectives in this area include reducing the end-to-end latency and interaction overhead between the processor and network interface. Performance optimizations include tighter integration of the network interface hardware with computing elements, such as the processor core.

IN THIS ISSUE

This special issue brings together four articles that address the key issues in network interface design across a wide range of cost-performance.

First, Simon Walton, Anne Hutton, and Joe Touch ("High-Speed Data Paths in Host-Based Routers") examine the use of commodity workstations and enhanced network interface cards to route Internet Protocol (IP) packets. The authors show that routing performance can be boosted by directly transferring the packet payload between the source and destination network interfaces, instead of staging it through the host memory.

Raoul A.F. Bhoedjang, Tim Ruhl, and Henri E. Bal ("User-Level Network Interface Protocols") present a tutorial on network software design for SANs. This article provides perspective on the wide range of high-speed protocols and interfaces built on Myrinet's network hardware as well as insights into critical design issues, such as data transfer, address translation, protection, and control transfer.

Thorsten von Eicken and Werner Vogels' "Evolution of the Virtual Interface Architecture" describes how several university research projects on SAN network interfaces helped shape the industry-standard NI specification called *Virtual Interface Architecture*. The VIA standard is supported by several hundred companies and is an emerging standard for cluster or system-area networks being jointly promoted by Intel, Compaq, and Microsoft.

Finally, Whay Sing Lee et al.'s "An Efficient, Protected Message Interface" shows how a tightly coupled parallel system, the M-Machine, can integrate a network interface with a processor to provide extremely low-latency communication. In the MIT M-Machine, the network interface sits directly next to the processor instead of on the I/O bus. This allows the M-Machine processors to launch messages directly into the network from the processor registers and integrate communication events with processor scheduling mechanisms.

We also refer readers to Shubhendu Mukherjee and Mark Hill's "Making Network Interfaces Less Peripheral" (*Computer*, Oct. 1998)—submitted and reviewed independently of this special issue—which argues that SAN network interfaces should appear to processors more like memory than a disk interface.

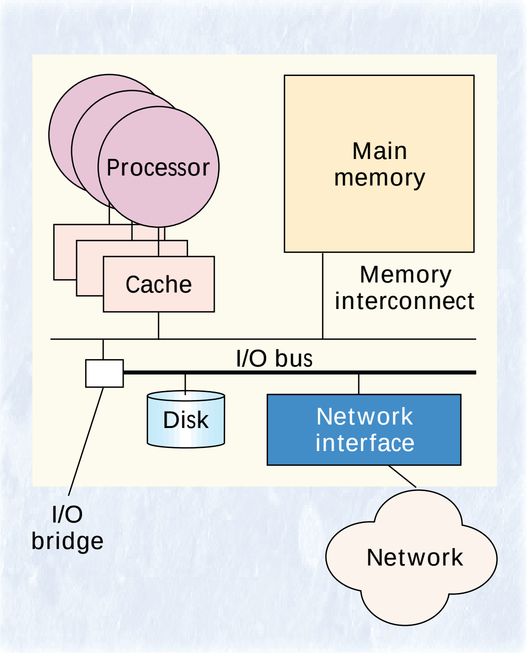

Figure 1. Architecture of a workstation node with a network interface attached to the I/O Bus.

Network interface design is a research topic of longstanding importance, with a huge body of literature and strong conference support. (See the "For More Information" sidebar.) Comprehensive coverage is impractical in a single special issue. Nevertheless, we hope that this issue provides a

For More Information

Hot Interconnects: A Symposium on High-Performance Interconnects. Sponsored by the IEEE Computer Society.

ACM SIGCOMM: Applications, Technologies, Architectures, and Protocols for Computer Communications. Sponsored by the ACM.

International Symposium on Computer Architecture (ISCA). Sponsored by the ACM and the IEEE.

International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS). Sponsored by the ACM.

International Symposium on High-Performance Computer Architecture (HPCA). Sponsored by the IEEE Computer Society.

glimpse of problems and challenges that lie ahead in the design of high-performance network interfaces. ❖

Acknowledgments

We gratefully acknowledge the efforts of both authors and reviewers for submitting excellent papers and reviews. We received 28 manuscripts and obtained a total of 96 reviews.

Andrew A. Chien is the SAIC Chair Professor in the Department of Computer Science and Engineering at the University of California, San Diego. His research involves networks, network interfaces, and the interaction of communication and computation in high-performance systems. Chien received a BS in electrical engineering and an MS and a PhD in computer science from the Massachusetts Institute of Technology. Chien is a recipient of a 1994 National Science Foundation Young Investigator Award.

Mark D. Hill is a professor and Romnes Fellow in the computer sciences department and the Electrical and Computer Engineering Department at the University of Wisconsin-Madison. He also codirects the Wisconsin Wind Tunnel parallel-computing project. His current research interests include memory systems of shared memory multiprocessors and high-performance uniprocessors. Hill received a BSE from the University of Michigan, Ann Arbor, and an MS and a PhD in computer engineering from the University of California, Berkeley.

Shubhendu S. Mukherjee is a senior hardware engineer on the Alpha Architecture team at Compaq Computer Corp. His research interests include network interfaces for system area networks, coherence protocols for shared memory multiprocessors, and microarchitectures for high-performance uniprocessors and multiprocessors. Mukherjee received a BTech from the Indian Institute of Technology, Kanpur, and an MS and a PhD from the University of Wisconsin-Madison.

Contact the guest editors at achien@cs.ucsd.edu, markhill@cs.wisc.edu, and shubu@muhthr.hlo.dec.com.